High Quality Holographic Stereograms Generation using four RGBD Images

SARAH FACHADA, DANIELE BONATTO, AND GAUTHIER LAFRUIT

Laboratories of Image Synthesis and Analysis, Université Libre de Bruxelles, Brussels, Belgium
gauthier.lafruit@ulb.ac.be

Abstract: In order to computer generate high quality holographic stereograms, a huge number of images must be provided: several hundreds for only horizontal parallax and the square of this number for full parallax. In this paper, we propose to drastically reduce this number to four input images with depth maps (or equivalently, four groups of neighbouring images used to compute a depth map) in any pose, for creating the missing images with depth image-based rendering. We evaluate the view synthesis method objectively before providing visual results of the corresponding holographic stereograms. This method outperforms shearlet-based approaches in objective view synthesis quality metrics and in the number of required input images (7×7).

© 2021 Optical Society of America

1. Introduction

During the past recent years, holography has spread in our daily life [1–3]. Holographic stereograms represent a light field of the scene [4], and come either in Horizontal Parallax Only (HPO) format, or in full parallax (horizontal and vertical parallax). However, the viewpoints must be perfectly aligned and in the order of several hundreds of them for HPO, demanding a professional image acquisition setup ie. a sliding camera on a rail with acquisitions every millimeter at a distance 1m to the scene [5] (natural scenes) or knowledge in 3D rendering (3D modelled scenes). The purpose of this paper is to reduce the number of input images necessary to create a high quality holographic stereogram of a natural scene without explicit 3D reconstruction or complex and expensive scanners/lasers. To reach high quality results, our Depth Image-Based Rendering (DIBR) pipeline requires only four images with corresponding depth maps (RGBD images) to recreate the dense light field needed for the hologram generation. When the depth maps are estimated, four supplementary viewpoints acquisition per input view are used for accurate stereo-matching with DERS software [6], a tool recommended in a new standard for immersion, MPEG-I (see subsection 2.2).

1.1. Image acquisitions for holography

An excellent overview of the various acquisition methods for digital holography can be found in [7] both for full holograms and holographic stereograms. While the creation of full holograms calls for a 3D knowledge of the scene (direct acquisition [2, 5], point cloud [8], mesh [9, 10]), holographic stereograms are related to its light field [4] representation (direct photos acquisition [3, 11], image based rendering [12–14]). Indeed, a full hologram recreates the 3D points of an object directly in space by forming interference patterns, contrary to the holographic stereogram which is composed of hogels (holographic equivalent of a pixel) emitting light rays with the appropriate color and direction from its surface.

It is possible to generate holograms with the point cloud given by a RGBD image (eg. obtained by stereo matching [15] or depth sensing device [16]), nevertheless, the price to pay is that the field of view then remains very limited around the viewpoint position.

Hologram computation using synthesized views instead of only captured images was first introduced by Katz et al. [17] to avoid complex acquisitions using 1D micro arrays [18] or translation stages [11]. However, Katz et al. use a basic technique for estimating novel views:

after the detection of the vertical lines using a convolution kernel over two views to compute the sparse correspondence between the views, they estimate a dense displacement map. While this approach is suited for small objects with few occlusions, it cannot be easily generalized to complex scenes. The second drawback of the technique is that only views in-between two known views can be synthesized, but no extrapolation is possible outside the area delimited by the captured views.

In [14], Gilles et al. used a Multiview-plus-Depth (MVD) approach, similar to this paper. Nevertheless, they use a point-cloud to render the scene from novel viewpoints. The accuracy of 2D image based point clouds is highly dependent on the number and configuration of the input images. In general, point-cloud based methods work well for unstructured objects, however, the lack of topology calls for splatting techniques [19] to fill the holes in-between the points. In their paper, Gilles et al. used 9×9 cameras with a spacing of 0.3mm. They used synthetic scenes and cameras with perfect rendered depth maps, but do not provide results on natural scenes.

Another category of techniques to retrieve missing information in a light field are the shearlet-based methods [20]. Those techniques do not explicitly recover the geometry of the scene [21] but can reconstruct a light field from a sparse acquisition ending up with satisfying HPO results from as few as 17×1 viewpoints, eg. PSNR of 39 to 41dB [22]. However those results are obtained for view interpolation between very close input images. In practice, for holography, an extrapolation on a wider range is needed, and the PSNR results are around 23db for an full parallax input camera array of 7×7 viewpoints [23].

1.2. Depth image-based rendering

DIBR [24] is a rendering technique consisting in warping two or more input images to novel viewpoints according to their depth information: foreground pixels have a higher disparity than background pixels, thus they get more motion parallax. Hence DIBR assumes a prior step of camera parameters and depth estimation. The depth maps can be estimated using stereo-matching [25, 26], learning [27], active camera devices, structured lights [28]. A recent study comparing different view interpolation methods [29] concludes that traditional DIBR, simply based on geometric information of the scene, remains competitive against more recent advances, including learning-based methods.

This technique is hence suited to render natural scenes without the step of reconstructing and rendering 3D models with structure from motion [30–32]. Creation of a 3D mesh from photographic acquisitions is called photogrammetry. Even if it provides useful topological information, it does not render as photo-realistic images as DIBR (see section 3). DIBR is also more resistant to artifacts from errors in depth estimation than shearlet-based approaches [7]. However, it can suffer from specific artifacts [33], due to the camera calibration, depth estimation or missing information when the number of images is limited. Errors in the depth maps and calibration cause ghosting artifacts (blurry images, object appearing in double) [34]. Navigation towards the scene (step-in) or zooming creates dilation holes with the appearance of "cracks" [35], but can be solved by splatting [35, 36], superpixels [37] or implicit meshing [38]. When an object is hidden in some input viewpoints but should appear in the rendered ones, the missing information creates a disocclusion. Disocclusion problems are handled using inpainting methods [39, 40] or selecting an increased number of input views with wide baselines [38, 41].

1.3. Overview

Our pipeline (figure 1) covers the steps from scene acquisition to holographic stereogram creation and provides a visual overview of this paper. We render the images needed for the hologram generation using DIBR, after a first step of calibration and depth estimation for the case of natural scenes. DIBR relies on a very limited number of input images (at the minimum one image and its corresponding depth map) and outputs photo-realistic new viewpoints. In this paper, we compare

Fig. 1. Proposed pipeline from images acquisition to holographic stereogram, with view synthesis quality evaluation. Left: natural dataset ToysTable. Right: Synthetic dataset Classroom. Step 1 corresponds to section 2.1. Step a.2 is explained in section 2.2. Section 3 details step 3 and section 2.3 details step 4. Step 5 corresponds to section 4. Step 6 is the final result, shown in section 5.

DIBR hologram creation with four and eight input RGBD images, proving that four viewpoints is sufficient. Given input images, and their associated depth maps, we render all the images required to create a digital hologram with imperceptible artifacts. Two or more input images are required to generate a depth map using stereo-matching (section 2.2). Our view synthesis software (detailed in section 2.3) obtains high PNSR scores for synthetic and natural content, hence reaching the high quality views used to create the hologram (section 3). We show that even with estimated depth maps, high quality results can be obtained. The details of holographic stereogram creations are provided in section 4. In section 5, we show holographic stereograms computed with our method on both natural and synthetic datasets and compare the results in four final holographic stereograms: one is made with DIBR with four input images from a synthetic dataset, the three other are made from a natural dataset (DIBR with 4 and 8 input images and depth maps, and finally, views rendered from a photogrammetric mesh). Although, our manufactured stereograms are HPO, our technique applies identically to full parallax holography.

2. Holographic content creation

We synthesize the views needed to create the holograms with DIBR. We use the process from acquisition to rendering initially designed by MPEG-I (where the "I" stands for immersive; MPEG-I is the standardization activity for the compression of immersive content) for content compression [38, 42, 43]. At the time of writing, a holographic stereogram was printed with up to 768 images, that we synthesized in their entirety. The Classroom dataset rendering, ToysTable depth maps, and view synthesis have been performed on a computer equipped with an Intel Xeon E5-2680@2.7GHz CPU and a NVIDIA GTX 1080TI GPU.

2.1. Datasets creation

In this work, we created holographic stereograms with natural and synthetic content (respectively rows a and b in figure 1). Two datasets have been used. The first, Classroom [44, 45], is a synthetic 3D scene available as a demo for Blender [46], a 3D editing software. The second, ToysTable [47] is a natural scene, providing a dense sampling of the light field, acquired by a translation stage with 3 degrees of freedom (figure 3). ToysTable was previously named ULB Unicorn in [47], before being renamed for MPEG-I's standardisation purposes.

Classroom The Classroom dataset [44] is a synthetic scene made for 3D modelling with the Blender editing software. It has been used as a test scene in MPEG-I exploration experiments [45]. In this scene, we use the rendered depth maps, thus creating the conditions to avoid the artifacts due to their estimation (figure 2).

ToysTable The ToysTable dataset is a dense multiview capture of a scene made of toys put on a table. It is composed of a scene sampled every millimeter thanks to a 3 degrees of freedom robotic setup, shown in figure 3, though to create the stereograms, only four to eight views were used; all other views were used as reference in the quality evaluation (see section 3). The cameras are disposed in three 2D arrays at various distances from the scene (55cm, 85cm, 105cm) and on a path line towards the scene. It was acquired using a Kinect v2. The dataset is provided with Kinect depth maps, estimated depth maps, and a photogrammetric point cloud and mesh created with the Reality Capture software [48].

2.2. Depth estimation

As DIBR requires depth information to render new viewpoints, a dataset must be accompanied by its corresponding depth maps. For the synthetic dataset Classroom, we rendered depth maps corresponding to the ray-traced images. For the natural dataset Toystable, we used MPEG-I's

Fig. 2. The four input viewpoints used to generate the 768 views of the holographic stereogram Classroom.

Fig. 3. Three degrees of freedom robotic acquisition system used for the ToysTable dataset.

depth estimation software DERS [6, 49, 50]. DERS is based on plane sweeping between two to five stereo input images in various configurations in a first pass. It refines the results in a second pass with a graph-cut approach to obtain high quality depth maps [51]. Though we had acquired the ToysTable scene with a Kinect v2 to obtain depth maps, an objective comparison shows that estimated depth maps provide more accurate results (up to 5dB of PSNR) than those provided from the acquisition with a depth sensing device [47]. Stereo-matching algorithms ask for more than one image to estimate a depth map even if ultimately only one input is used for view synthesis, while a depth sensing device does not require any additional view. In our case, even if one of the ToysTable holograms has been created using four input images and depth maps, adding the depth maps estimation increases the total number of used acquisition to sixteen for the four RGBD input hologram and twenty for the eight RGBD hologram (see figure 1(a.2): plain red for the four RGBD input, plus yellow for the eight RGBD input). Nevertheless, theoretically if the images used as input for DIBR are sufficiently close, the stereo-matching can be performed without additional images.

2.3. *Rendering novel views*

Novel views are computed using the Reference View Synthesis (RVS) software [38, 42, 52] developed in the context of MPEG-I replacing the anchor reference software View Synthesis Reference Software (VSRS) [53], that did not reach the state-of-the-art results anymore. A demo of RVS can be found at <https://lisaserver.ulb.ac.be/rvs>. RVS has been successively improved [52, 54, 55] bringing DIBR up to a high level of quality. First, its implementation relies on implicit meshing of the pixels from the input views, allowing the rendering of distant viewpoints in any direction [54], contrary to VSRS or shearlet-transform methods, which are limited to interpolation and see their quality quickly decrease in case of extrapolation. Second, the input images and depth maps can be delivered in any configuration and number, contrary to the classical left-right stereo configuration. Finally, its GPU implementation carries the view synthesis to real-time performance [52] (used for the view synthesis in this paper). Previous experiments [47] show that four input images with corresponding depth maps were sufficient to reach high quality view synthesis results. Only four views were used to synthesize the hundreds of images needed for the hologram printing (figure 1(a,b.4)). Another hologram, computed from eight inputs images and depth maps has been created to confirm this result in the holographic application case.

3. **View synthesis results**

In this section, we show objective results of the view synthesis measured in PSNR (on color images) and MS-SSIM [56] (on gray-scale images). PSNR (measured in dB) is a quality metric comparing two images pixel-wise, while MS-SSIM (measured between 0 and 1) focuses on their structural similarity. Higher values correspond to better results. Admittedly PSNR is sensible to slight pixel-shifts, while MS-SSIM quickly reaches values close to 1. Finding meaningful objective quality metrics for view synthesis is a whole domain of exploration, which led to the creation of an activity in the MPEG-I group. For the synthetic dataset, we compare ray-traced views taken as ground truth to the views rendered with DIBR (see figure 1(b.2,3)). For the natural dataset, we compare our DIBR rendered views to our acquisitions taken as ground truth (see figure 1(a.3)).

We reach a PSNR of 24.9dB for the synthetic dataset Classroom with images computed from four input viewpoints and 23.50dB (resp. 23.96dB) for the natural dataset ToysTable with images computed from four (resp. eight) input viewpoints. For comparison, shearlet-based approaches [23] reach PSNRs around 23dB for 7×7 parallel input images (three times more than the total of 16 acquisitions).

Fig. 4. (a) PSNR and (b) MS-SSIM of the synthesised views used to create the holographic stereogram of Classroom dataset. Dark blue line: full image comparison. Orange line: comparison with masked disocclusions.

3.1. Synthetic dataset

A synthetic dataset allows to render ground truth images from any viewpoint directly by ray-tracing in Blender. To objectively evaluate our view synthesis, we have chosen to compare the 768 DIBR views used to create the holographic stereogram to their ray-traced ground truth (see figure 1(b.2)). The 768 views have been generated on a circular arc of 120° from only four input images widely spaced and focusing on the central part of the scene (see figure 2).

Figure 4 shows the PSNR (up to 24.9dB) and MS-SSIM (up to 0.981) of the images used to generate our holographic stereogram. The views on the extremities of the arc reach a lower score because of the numerous disocclusions and parts of the scene which do not appear in any input viewpoint of figure 2. Those areas are colored in black. Another color choice could have modified the score results. A drop in PSNR values appears between views 420-450, due to specular reflections of the windows light observed at those viewpoints. We also provide the scores with masked occlusions to evaluate the quality of the view synthesis without unfilled disocclusions (PSNR up to 26.7dB and SSIM up to 0.993).

Visual results are shown in figure 5. Slight differences can be observed due to non-perfectly diffuse parts of the scene (table top, window of the door).

3.2. Natural dataset

	PSNR (dB)	PSNR (dB)	MS-SSIM	MS-SSIM
	4 input	8 input	4 input	8 input
MIN	13.13	15.71	0.732886	0.843867
MAX	23.50	23.96	0.988463	0.989082
MEAN	19.52	21.60	0.947457	0.970915

Table 1. PSNR and MS-SSIM for view synthesis in ToysTable dataset with 4 and 8 input viewpoints.

In the ToysTable dataset, although the Kinect depth maps are available, we used the depth maps estimated with DERS to reach higher accuracy [47], at the cost of increasing the number of acquisition required (for four (resp. eight) input RGBD images with depth maps, sixteen (resp. twenty) acquisitions were made). Contrary to our synthetic dataset, only a dense set of

Fig. 5. (a), (c) DIBR-rendered viewpoints used to compute the hologram. Disocclusions are displayed in black. (b), (d) Corresponding viewpoints rendered from the 3D model for quality assessment.

Fig. 6. DIBR-rendered views used to create holographic stereograms with (a) 4 input viewpoints, (b) 8 input viewpoints. No ground truth image is available in this dataset (this natural dataset is composed of parallel cameras grouped in a small dense area in front of the scene).

Fig. 7. Rendered objects with photogrammetry using several hundreds of images acquired to create the 3D model by photogrammetry.) (a),(c) and DIBR (from four images). (b),(d) appearing in our holographic stereograms.

parallel camera images is available. It does not include the viewpoints of the hologram, hence we measured the PSNR and MS-SSIM for view synthesis of the available viewpoints (see table 1). Some rendered views that have been used to create the hologram are shown in figure 6.

Eventually, we propose a visual comparison between photogrammetric rendering of some objects of the dataset and the same objects rendered with DIBR in figure 7, as we also created a hologram from the photogrammetric mesh. Despite the hundreds of views required to obtain a textured mesh of an object, 3D reconstruction is sensible to fine material (here fluffy textures) and cannot render realistically the geometry and aspect of those objects. By the use of only implicit geometry and warping of images, DIBR allows a much more convincing appearance (see figure 7).

4. From synthesized images to holographic stereograms

We outsourced the printing of several HPO holographic stereograms representing the ToysTable and Classroom datasets with the printing methods described in the following paragraphs. Although the holographers use different techniques for the holographic stereogram creation, the input is similar: hundreds of synthetic views generated using four RGBD images which are converted to

Fig. 8. The perspective images rendered with DIBR are converted to hogels with the method described in [57, 58]. Left: configuration for the ToysTable dataset. Right: configuration for the Classroom dataset.

hogels by the holographers [58] before the stereogram is engraved (see figures 1(a,b.5) and 8). The first holographer [1] requires the views to be sampled on a horizontal line parallel to the scene with their orientation toward a central point of the scene and with an angle of -35° to 35° from it (ToysTable dataset). The second technique [2] (CHIMERA), described in [59], requires an image sequence on a 120° circular arc around the scene (Classroom dataset). The video acquisition is replaced by the DIBR images. The holographer converts the views to hogels with the method described in [57]. A smaller hogel size corresponds to a better resolution and visual sharpness.

ToysTable Three holograms were manufactured by Swetter [1]. Two were created with DIBR using 400 synthesized views from 4 and 8 input cameras (see figure 9(a)) and one with 500 views rendered from the photogrammetric model (see figure 9(b)). The holograms were engraved with a hogel size of 500m in a frame of $26\text{cm} \times 21\text{cm}$.

Classroom The Classroom hologram has been printed using a better hogel resolution, more output views, and only four input images with their depth for the RVS view synthesis. They were engraved with the CHIMERA method [2, 59]. In total, 768 views were synthesized from 4 input viewpoints and their rendered depth maps. The corresponding hologram has a hogel size of 250m for a total hologram size of $40\text{cm} \times 22\text{cm}$. The hologram and synthesized views (close-ups) are shown in figure 9(c).

5. Holography results

Our first comparison is between the two holograms of ToysTable created with 400 views rendered with DIBR. The first is computed using RVS with only four input viewpoints and the second eight and their depth maps. One could expect that it is preferable to use eight input viewpoints in accordance to table 1, but visually very few differences are perceptible in the rendered images (figure 6 and no difference is perceptible in the holograms, probably due to the 500m hogel resolution causing some blurriness (figure 9(a)).

We also tested the possibility of generating a hologram using photogrammetry. It is interesting to note that, despite the unrealistic appearance of models rendered with photogrammetry, cf section 3.2, the low-resolution stereogram rendered with this method does not suffer too much from the same disturbing appearance (figure 9(b)), except a hole in the color chart which is a photogrammetric artifact. However, at higher resolutions, errors in the mesh will appear more clearly.

Finally, given the conclusions made from the two previous experiments — a higher hogel

(a)

(b)

(c)

Fig. 9. (a) Comparison of DIBR holograms (top) synthesized with (left) 4 input views (see Visualization 1) and (right) 8 input views (see Visualization 2) (400 views, hogels of 500m). (bottom) images rendered with RVS to compute the hologram. (b) Comparison of holograms created from (left) a photogrammetric mesh (see Visualization 3) (500 synthesized output views, hogels of 500m) and (right) DIBR (400 synthesized output views, 4 input images, hogels of 500m). (c) Holographic stereogram computed from the synthetic dataset Classroom looked at from two viewpoints (see Visualization 4) (768 synthesized output views, 4 input images, hogels of 250m) (left) with close synthesized views (right). (Paper clips to hold the hologram are visible at the top of the left pictures.)

Fig. 10. Recording a hologram using a camera fixed on a circular rail to obtain a smooth curved movement.

resolution is preferable and four input viewpoints are sufficient — we created the holographic stereogram of the Classroom dataset (figure 9(c)). The hogels are not visible contrary to the other holograms, but the nature of the scene, presenting many occluding objects made a few disocclusion artifacts appear (principally behind the desk and the coat hanger). Nevertheless, the level of detail is more than satisfying.

Videos of all our holograms are provided in the supplementary materials (Visualizations 1 to 4). They have been captured using a camera mounted on a curved rail as shown in figure 10.

6. Conclusion

A DIBR approach is a promising method for holographic stereograms generation for a large variety of natural and synthetic datasets. With a couple of examples, we have shown that good PSNR and MS-SSIM measures can be expected, as well as good visual results in holography. DIBR enables the use of relatively large distances between the cameras in practical contexts of natural scenery and requires a very limited number of image acquisitions (in any pose): as little as four viewpoints with depth maps are sufficient for good results in view synthesis while avoiding classical interpolation artifacts. In future work, the quality of the holographic stereograms will be tested using a keratometer. Eyes' accommodation and vergence will be measured in complement to a subjective evaluation to confirm our results. One may indeed expect that for a sufficiently high quality of the hologram (resolution, hogel size and content), the eyes should focus and converge on the virtual 3D objects and not on the hologram plane itself [60–62].

Funding

This work was supported by: 1) Innoviris, the Brussels Institute for Research and Innovation, Belgium, under contract n° 2015 R39c, 3DLicorneA. 2) the Fonds de la Recherche Scientifique-FNRS, Belgium, under Grant n° 33679514, ColibriH. 3) the EU project n° 951989 on Interactive

Technologies, H2020-ICT-2019-3, Hovitron. Sarah Fachada is a Research Fellow of the Fonds de la Recherche Scientifique - FNRS, Belgium.

Acknowledgments

The multiview dataset ToysTable was created with the support of Innoviris under the project 3DLicorneA. FNRS supported the RVS research with the project ColibriH. The EU project Hovitron allowed privileged access to advanced MPEG-I software tools. Sarah Fachada is a Research Fellow of the Fonds de la Recherche Scientifique - FNRS, Belgium.

Portions of this work were presented at OSA Imaging and Applied Optics Congress in 2020, Computer Generated Holography with Depth-based View Synthesis [63].

Disclosures

The authors declare no conflicts of interests.

References

1. E. Swetter, "3d hologrammen," (2019). <http://www.3-dhologrammen.com/>.
2. Y. Gentet, "Ultimate holography," (2019). www.ultimate-holography.com.
3. R. Munday, "A holographic portrait of Queen Elizabeth II," 7th Int. Symp. on Disp. Hologr. (2006).
4. M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96*, (ACM Press, 1996), pp. 31–42.
5. G. Saxby, *Practical Holography* (CRC Press, 2003).
6. T. Senoh, K. Yamamoto, N. Tetsutani, and H. Yasuda, "Enhanced DERS for quad reference views (eDERS) [M41955]," ISO/IEC JTC1/SC29/WG11 (2018).
7. J.-H. Park, "Recent progress in computer-generated holography for three-dimensional scenes," *J. Inf. Disp.* **18**, 1–12 (2017).
8. Y. Zhao, S. Alam, S.-H. Jeon, and N. Kim, "Fast calculation method for full-color computer-generated hologram with real objects captured by a depth camera," *Electron. Imaging* pp. 250–1–250–6 (2018).
9. Y. Zhao, K.-C. Kwon, Y.-I. Piao, Y.-T. Lim, S.-H. Jeon, and N. Kim, "Computer generated holograms of real object from depth camera using polygon-based method," in *Digital Holography and Three-Dimensional Imaging*, (Optical Society of America, 2016), pp. DW5I–8.
10. K. Wakunami, R. Oi, T. Senoh, H. Sasaki, Y. Ichihashi, and K. Yamamoto, "Wavefront printing technique with overlapping approach toward high definition holographic image reconstruction," in *Three-Dimensional Imaging, Visualization, and Display 2016*, vol. 9867 (International Society for Optics and Photonics, Baltimore, Maryland, United States, 2016), p. 98670J.
11. J. Jurik, T. Burnett, M. Klug, and P. Debevec, "Geometry-corrected light field rendering for creating a holographic stereogram," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (IEEE, Providence, RI, USA, 2012), pp. 9–13.
12. T. Senoh, K. Yamamoto, R. Oi, T. Mishina, and M. Okui, "Computer Generated Electronic Holography of Natural Scene from 2D Multi-view Images and Depth Map," in *2008 Second International Symposium on Universal Communication*, (IEEE, Osaka, Japan, 2008), pp. 126–133.
13. M. Yamaguchi, "Ray-based and wavefront-based holographic displays for high-density light-field reproduction," in *Three-Dimensional Imaging, Visualization, and Display 2011*, vol. 8043 (International Society for Optics and Photonics, 2011), p. 804306.
14. A. Gilles, P. Gioia, R. Cozot, and L. Morin, "Computer generated hologram from Multiview-plus-Depth data considering specular reflections," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, (IEEE, Seattle, WA, USA, 2016), pp. 1–6.
15. S.-C. Kim, D.-C. Hwang, D.-H. Lee, and E.-S. Kim, "Computer-generated holograms of a real three-dimensional object based on stereoscopic video images," *Appl. optics* **45**, 5669–5676 (2006).
16. F. Yang, Y. Wang, R. Mouthaan, and T. Wilkinson, "Holographic rendering of a real-world scene captured with a low-cost rgb-d camera," in *Digital Holography and Three-Dimensional Imaging*, (Optical Society of America, 2020), pp. HF4D–3.
17. B. Katz, N. T. Shaked, and J. Rosen, "Synthesizing computer generated holograms with reduced number of perspective projections," *Opt. Express* **15**, 13250 (2007).
18. Y. Sando, M. Itoh, and T. Yatagai, "Full-color computer-generated holograms using 3-D Fourier spectra," *Opt. Express* **12**, 6246 (2004).
19. M. Zwicker, H. Pfister, J. van Baar, and M. Gross, "Surface splatting," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01*, (ACM Press, Not Known, 2001), pp. 371–378.
20. G. Kutyniok, J. Lemvig, and W.-Q. Lim, "Shearlets and Optimally Sparse Approximations," in *Shearlets*, G. Kutyniok and D. Labate, eds. (Birkhäuser Boston, Boston, 2012), pp. 145–197.

21. F. Colonna, G. Easley, K. Guo, and D. Labate, "Radon transform inversion using the shearlet representation," *Appl. Comput. Harmon. Analysis* **29**, 232–250 (2010).
22. S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Image based rendering technique via sparse representation in shearlet domain," in *2015 IEEE International Conference on Image Processing (ICIP)*, (IEEE, Quebec City, QC, Canada, 2015), pp. 1379–1383.
23. E. Sahin, S. Vagharshakyan, J. Makinen, R. Bregovic, and A. Gotchev, "Shearlet-domain light field reconstruction for holographic stereogram generation," in *2016 IEEE International Conference on Image Processing (ICIP)*, (IEEE, Phoenix, AZ, USA, 2016), pp. 1479–1483.
24. C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291 (International Society for Optics and Photonics, 2004), pp. 93–105.
25. J. Xu, Q. Yang, and Z. Feng, "Occlusion-Aware Stereo Matching," *Int. J. Comput. Vis.* **120**, 256–271 (2016). Publisher: Springer.
26. O. Stankiewicz, G. Lafruit, and M. Domański, "Multiview video: Acquisition, processing, compression, and virtual view rendering," in *Academic Press Library in Signal Processing, Volume 6*, (Elsevier, 2018), pp. 3–74.
27. Y. Li, L. Zhang, Q. Wang, and G. Lafruit, "MANet: Multi-scale aggregated network for light field depth estimation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (IEEE, 2020), pp. 1998–2002.
28. Z. Zhang, "Microsoft Kinect Sensor and Its Effect," *IEEE MultiMedia* **19**, 4–10 (2012).
29. D. Yue, M. S. Khan Gul, M. Batz, J. Keinert, and R. Mantiuk, "A Benchmark of Light Field View Interpolation Methods," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, (IEEE, London, United Kingdom, 2020).
30. M. Brown and D. Lowe, "Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets," in *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, (IEEE, Ottawa, ON, Canada, 2005), pp. 56–63.
31. N. Snavely, S. M. Seitz, and R. Szeliski, "Photo Tourism: Exploring Photo Collections in 3D," in *ACM Siggraph 2006 Papers*, (2006), pp. 835–846.
32. S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building Rome in a day," *Commun. ACM* **54**, 105 (2011).
33. W. Sun, L. Xu, O. C. Au, S. H. Chui, and C. W. Kwok, "An overview of free view-point depth-image-based rendering (DIBR)," in *APSIPA Annual Summit and Conference*, (2010), pp. 1023–1030.
34. M. Uyttendaele, A. Eden, and R. Szeliski, "Eliminating ghosting and exposure artifacts in image mosaics," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2 (IEEE Comput. Soc, Kauai, HI, USA, 2001), pp. II–509–II–516.
35. Y. Mao, G. Cheung, and Y. Ji, "Image interpolation for DIBR viewsynthesis using graph fourier transform," in *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, (IEEE, Budapest, Hungary, 2014), pp. 1–4.
36. T. Senoh, K. Yamamoto, N. Tetsutani, and H. Yasuda, "VSRS improvement," in *ISO/IEC JTC1/SC29/WG11, M38979, 116th MPEG Meeting*, (2016).
37. M. P. Tehrani, T. Tezuka, K. Suzuki, K. Takahashi, and T. Fujii, "Free-viewpoint image synthesis using superpixel segmentation," *APSIPA Transactions on Signal Inf. Process.* **6** (2017).
38. S. Fachada, D. Bonatto, A. Schenkel, and G. Lafruit, "Depth Image-Based View Synthesis with Multiple Reference Views for Virtual Reality," in *2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, (IEEE, Helsinki, 2018), pp. 1–4.
39. A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," *J. Graph. Tools* **9**, 23–34 (2004).
40. O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image Inpainting: A Review," *Neural Process. Lett.* **51**, 2007–2028 (2020).
41. S. Li, C. Zhu, and M.-T. Sun, "Hole Filling with Multiple Reference Views in DIBR View Synthesis," *IEEE Transactions on Multimed.* **20**, 1948–1959 (2018). ArXiv: 1802.03079.
42. B. Kroon, "Reference View Synthesizer (RVS) manual [N18068]," ISO/IEC JTC1/SC29/WG11 p. 19 (2018). Edition: MPEG2013/W18068 Place: Macau SAR, China.
43. G. Lafruit, D. Bonatto, C. Tulvan, M. Preda, and L. Yu, "Understanding MPEG-I Coding Standardization in Immersive VR/AR Applications," *SMPTE Motion Imaging J.* **128**, 33–39 (2019).
44. C. Seux, *Blender - Classroom demo scene* (Blender Foundation, 2017).
45. B. Kroon, "3DoF+ test sequence ClassroomVideo [M42415]," ISO/IEC JTC1/SC29/WG11 (2018). Edition: MPEG2018/M42415 Place: San Diego, CA, US.
46. Blender Online Community, "Blender - a 3D modelling and rendering package," (2019). <http://www.blender.org>.
47. A. Schenkel, D. Bonatto, S. Fachada, H.-L. Guillaume, and G. Lafruit, "Natural Scenes Datasets for Exploration In 6DOF Navigation," in *2018 International Conference on 3D Immersion (IC3D)*, (IEEE, Brussels, Belgium, 2018), pp. 1–8.
48. "Capturing Reality," (2020). <http://www.capturingreality.com>.
49. S. Rogge, D. Bonatto, J. Sancho, R. Salvador, E. Juarez, A. Munteanu, and G. Lafruit, "MPEG-I Depth Estimation

- Reference Software,” in *2019 International Conference on 3D Immersion (IC3D)*, (IEEE, Brussels, Belgium, 2019), pp. 1–6.
50. K. Wegner, “Software Manual DERS,” ISO/IEC JTC1/SC29/WG11 p. 17 (2014).
  51. V. Kolmogorov and R. Zabih, “Multi-camera Scene Reconstruction via Graph Cuts,” in *Proceedings of the 7th European Conference on Computer Vision-Part III*, (Springer-Verlag, Berlin, Heidelberg, 2002), ECCV ’02, pp. 82–96.
  52. D. Bonatto, S. Fachada, and G. Lafruit, “RaViS: Real-time accelerated View Synthesizer for immersive video 6DoF VR,” in *IS&T Electronic Imaging 2020, The Engineering Reality of Virtual Reality Conference*, (Burlingame, USA, 2020).
  53. O. Stankiewicz, K. Wegner, M. Tanimoto, and M. Domański, “Enhanced View Synthesis Reference Software (VSRS) for Free-viewpoint television [M31520],” ISO/IEC JTC1/SC29/WG11 (2013). Edition: MPEG2013/M31520.
  54. S. Fachada, D. Bonatto, A. Schenkel, and G. Lafruit, “Free Navigation in Natural Scenery With DIBR: RVS and VSRS in MPEG-I Standardization,” in *2018 International Conference on 3D Immersion (IC3D)*, (Bruxelles, Belgium, 2018), pp. 1–6.
  55. S. Fachada, D. Bonatto, and G. Lafruit, “Xsliit cameras for Free Navigation with Depth Image-Based Rendering,” in *SPIE 2020, the International Society for Optics and Photonics, Photonics Europe*, (Online meeting, 2020).
  56. Z. Wang, E. Simoncelli, and A. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, (IEEE, Pacific Grove, CA, USA, 2003), pp. 1398–1402.
  57. J. Su, Q. Yuan, Y. Huang, X. Jiang, and X. Yan, “Method of single-step full parallax synthetic holographic stereogram printing based on effective perspective images’ segmentation and mosaicking,” *Opt. Express* **25**, 23523–23544 (2017).
  58. J. Su, X. Yan, Y. Huang, X. Jiang, Y. Chen, and T. Zhang, “Progress in the Synthetic Holographic Stereogram Printing Technique,” *Appl. Sci.* **8**, 851 (2018).
  59. Y. Gentet and P. Gentet, “CHIMERA, a new holoprinter technology combining low-power continuous lasers and fast printing,” *Appl. Opt.* **58**, G226 (2019).
  60. L. Ryan, *Vergence-accommodation conflicts and visual performance in stereoscopic 3-D imagery* (Bangor University (United Kingdom), 2015).
  61. Q. Y. J. Smithwick, J. Barabas, D. E. Smalley, and V. M. B. Jr., “Interactive holographic stereograms with accommodation cues,” in *Practical Holography XXIV: Materials and Applications*, vol. 7619 H. I. Bjelkhagen and R. K. Kostuk, eds., International Society for Optics and Photonics (SPIE, 2010), pp. 19 – 31.
  62. P. St Hilaire, P.-A. Blanche, C. Christenson, R. Voorakaranam, L. LaComb, B. Lynn, and N. Peyghambarian, “Are stereograms holograms? a human perception analysis of sampled perspective holography,” in *Journal of Physics: Conference Series*, vol. 415 (IOP Publishing, 2013), p. 012035.
  63. S. Fachada and G. Lafruit, “Computer Generated Holography with Depth-based View Synthesis,” in *Imaging and Applied Optics Congress (2020)*, (Optical Society of America, 2020), p. HF1D.7.